Лекция3

Распределение случайных величин.

Функция распределения и плотность распределения случайной величины.

Числовые характеристики случайной величины.

Математическое ожидание и дисперсия.

Пусть дискретная физическая величина X может принимать в результате опыта значения $X_1, X_2, \dots, X_n$. Отношение числа опытов m_i, в результате которых величина X принимает значение X_i, к общему числу проведенных опытов n называется частотой появления события $X = X_i$. Частота ${}^{m_i}/_{n}$ является случайной величиной и меняется в зависимости от количества проведенных опытов. Однако при большом количестве опытов (в пределе n $\rightarrow \infty$) она стабилизируется около некоторого значения p_i, называемого вероятностью события $X = X_i$ (статистическое определение):

$$p_i = P(X = X_i) \approx \frac{m_i}{n}$$

Сумма вероятностей реализации всех возможных значений случайной величины равна единице:

$$\sum_{i=1}^{n} p_i = 1$$

Дискретную случайную величину можно полностью задать вероятностным рядом, указав вероятность p_i для каждого значения X_i:

Распределения случайной величины

Законом распределения случайной величины называют любое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Вероятностный ряд является одним из видов законов распределения случайной величины.

Распределение непрерывной случайной величины нельзя задать вероятностным рядом, поскольку число значений, которое она может принимать, так велико, что для большинства из них вероятность принять эти значения равна нулю.

функция распределения

- Для непрерывных физических величин изучается вероятность того, что в результате опыта значение случайной величины попадет в некоторый интервал. Удобно пользоваться вероятностью события $X \leq x$, где x — произвольное действительное число. Эта вероятность

$$P(X \leq x) = F(x)$$

является функцией от x и называется функцией распределения (предельной функцией распределения, функцией распределения генеральной совокупности) случайной величины. В виде функции распределения можно задать распределение как непрерывной, так и дискретной случайной величины. $F(x)$ является неубывающей функцией, т.е. если

Плотность распределения

- Если $F(x)$ непрерывна и дифференцируема, то

$$f(x) = \frac{dF(x)}{dx}$$

- Задание $f(x)$ также полностью определяет случайную величину. Плотность распределения является неотрицательной функцией.
- Площадь, ограниченная осью x, прямыми $x = x_1$ и $x = x_1$ и кривой плотности распределения, равна вероятности того, что случайная величина примет значения из интервала $x_1 - x_2$

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)\,dx = F(x_2) - F(x_1)$$

Тогда

$$F(x) = P(-\infty \leq X \leq +\infty) = \int_{x_1}^{x_2} f(x)\,dx$$

Поскольку попадание случайной величины в интервал $-\infty < X < +\infty$ есть достоверное событие, то

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Вместо полного определения случайной величины в виде законов распределения вероятностей в прикладных задачах ее часто определяют при помощи числовых характеристик — чисел (вещественных), выражающих характерные особенности случайной величины, называемых моментами случайной величины. Наиболее часто в приложениях математической статистики используют:

математическое ожидание (характеристику положения значений случайной величины на числовой оси) дисперсию (или среднее квадратичное отклонение), определяющую характер разброса значений случайной величины.

Математическое ожидание

- Математическое ожидание (генеральное среднее) случайной величины (начальный момент первого порядка) принято обозначать $M[X]$, m или m_x. Оно определяется для дискретной и непрерывной случайной величины соответственно как

$$m = M[X] = \sum_{i=1}^{n} x_i p_i$$

$$M[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

$f(x)$ – дифференциальная функция распределения непрерывной случайной величины (или ее плотность вероятности)

Для случайных величин математическое ожидание является теоретической величиной, к которой приближается среднее значение x случайной величины X при большом количестве испытаний.

Дисперсия

- Дисперсией (вторым центральным моментом) случайной величины называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания, т. е.

$$D[X] = M\left[(X - M[X])^2\right]$$

Для дискретной и непрерывной случайных величин дисперсия определяется следующим образом соответственно:

$$D[X] = \sum_{i=1}^{n} (x_i - m_x)^2 p_i$$

$$D[X] = \int_{-\infty}^{+\infty} (x_i - m_x)^2 f(x) dx$$

Дисперсия является мерой рассеяния значений x около их математического ожидания.

Корень квадратный из второго центрального момента называется средним квадратичным отклонением (стандартным отклонением, или стандартом):

$$\sigma = \sqrt{D[X]}$$

При бесконечно большом числе наблюдений истинное значение измеряемой величины равно среднему арифметическому значению всех результатов наблюдений:

$$A = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} a_i = \bar{a} = \widehat{A}$$

Смысл выражения в следующем: точечной оценкой истинного значения измеряемой величины в случае нормального распределения наблюдений является среднее арифметическое по всем наблюдениям.

При конечной выборке *n* вместо математического ожидания можно получить только среднее значение:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

а вместо дисперсии среднеквадратическое отклонение:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Точечная оценка

Точечной оценкой называют оценку, которая определяется одним числом. При малом количестве обрабатываемых измерений *n* точечная оценка может значительно отличаться от оцениваемого параметра. Поэтому при небольшом объёме выборки необходимо рассмотреть надежность этой оценки, которую можно оценить ***неслучайным интервалом***, расположенным вокруг точечной оценки, в который результат измерения попадет с заданной доверительной вероятностью (обычно в измерениях называемой «надежностью» и обозначаемой α).

связь между истинным значением измеряемой величины A и её точечной оценкой

- Чем меньше число наблюдений *n*, тем больше величина $\bar{a}$ зависит от отдельных результатов наблюдений, но так как результаты наблюдений случайны, то среднее, найденное по конечному числу наблюдений, также будет случайной величиной. Обозначим $\Delta\bar{a}$ - отклонение точечной оценки от истинного значения: $\Delta\bar{a} = A - \widehat{A} = A - \bar{a}$. Отсюда видно, что из-за случайности средних случайными будут и отклонения $\Delta\bar{a}$. Однако, с увеличением числа усредняемых значений влияние величины каждого отдельного наблюдения на среднее становится меньше (действительно, с каждым новым значением к среднему прибавляется величина , где *n* - число наблюдений) и точечная оценка обретает так называемую ***статистическую устойчивость***, и отклонение оценки от истинного значения меньше зависит от отдельных наблюдений. По смыслу $\Delta\bar{a}$ - та погрешность, которую допускают, взяв вместо истинного значения его оценку – среднее арифметическое .

Установим связь между $\sigma_{\bar{a}}^2$ и дисперсией результатов наблюдений σ^2. Результаты наблюдений a_i - независимые случайные величины с дисперсией σ^2.

$$\sigma_{\bar{a}}^2 = \frac{\sigma^2}{n}$$

Дисперсия среднего из *n* наблюдений в *n* раз меньше дисперсии результаты наблюдения. Иными словами, если за результат измерения принять наблюдение, то разброс такой оценки будет иметь дисперсию σ^2, а если за оценку принять результат полученный усреднением *n* наблюдений, то эта оценка характеризуется в *n* раз меньшей дисперсией $\sigma_{\bar{a}}^2$.